

LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

HD28
.M414
no. 254-
67

A NOTE ON THE INTERPRETATION OF
FACTOR ANALYSIS
or
FACTOR ANALYSIS: WHAT GOOD IS IT?
J. Scott Armstrong
and Peer Soelberg

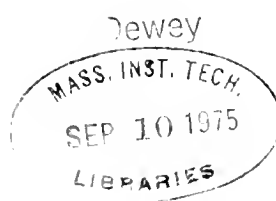
A NOTE ON THE INTERPRETATION OF
FACTOR ANALYSIS
or
FACTOR ANALYSIS: WHAT GOOD IS IT?
J. Scott Armstrong
and Peer Soelberg

254-67

© Copyright by Peer Soelberg and J. Scott Armstrong

This paper should not be reproduced in whole or in part,
by any process, without the authors' permission.

RECEIVED
SEP 10 1975
LIBRARIES



ABSTRACT

Results of a survey are examined by a common factor analytic method. The conclusion leads us to question the value of many recent papers that have used factor analysis as their chief research method. Recommendations are made for requirements that should be met by published articles employing factor analysis.

FACTOR ANALYSIS: WHAT GOOD IS IT?

We have been concerned with certain methodological problems of factor analysis. In order to get some feeling for the seriousness of the problems that concerned us, we decided to study the following specific application.

We asked 50 employees to rate their supervisors on 20 traits. The measuring instrument was a semantic differential ranging from +3 to -3. An example is provided in Figure 1.

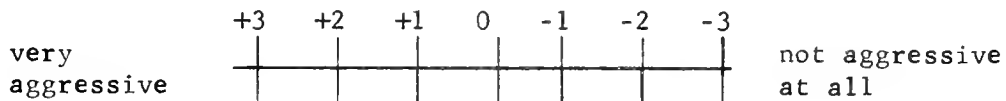


Figure 1

The employees were all selected randomly from different work groups. Analysis of the data showed that their responses to each trait were approximately normally distributed. The responses were then factor analyzed in order to determine whether underlying factors might be found that could summarize the results in a meaningful fashion. Since this was to be an exploratory study, we hoped to use any factors that emerged as our source of hypotheses for further work on employee perception of supervisory behavior.

Pearson product-moment correlations were computed between all trait variables. A principal components analysis was then carried out using unity in the major diagonal. Following the recommendations of Kaiser (1960), we selected all factors with eigenvalues greater than 1.0. These factors were then rotated according to the varimax criterion. All calculations were done with the California Biomedical 03M program at the MIT Computation Center.

Results

The results of the factor analysis seemed promising. The data did fall into a pattern. In fact, we were able to summarize 71 percent of the variance in the 20 traits with only 9 factors. The factors, using only loadings greater than 0.50, are shown in Table 1.

Insert Table 1 about here

From Table 1 we see that very impressive loadings were obtained. The factors also make good intuitive sense. Factors II, VI, and VIII came through clearly, with only one significant variable in each factor: Sincerity, Kindness, and Tactfulness, respectively. The other factors bear some discussion.

Factor I seems to be a measure of Fascism. It is closely related to the authoritarian personality described by Adorno, et.al. (1950). Supervisors who score highly on this factor are low in sensitivity, low in democratic attitudes, but high on responsibility. The relationship to responsibility also receives support from the "blind obedience" study by Milgram (1964).

Factor III is a measure of Social Distance. People who score high on this factor tend to be formal, somewhat unfair, and are low on humility. This makes sense since, where social distance is high, the supervisor does not gain sufficient information about his people and about particular problem situations; as a result his decisions will seem unfair. In addition, social distance can lead to the perception that the supervisor is making unilateral decisions, which by itself may appear unfair and presumptuous to employees in the American culture. The factor parallels Fiedler's description of social distance (1960).

Factor IV is a measure of Reliability. Supervisors high on this factor are trusted by their employees. The loading of humbleness on both Factor III and Factor IV is, however, not so easily explained. But it may be that the term "humble" is eliciting two types of responses. In Factor III it is a type of "aloofness", while in Factor IV it is more likely associated with a non-evaluative attitude.

Factor V might be called Docility. Supervisors scoring high on this factor are perceived as patient, unaggressive, and rather shallow. Hertzberg et.al. (1959) have shown that this type of person is more interested in the hygiene conditions of his work rather than in the task itself.

Factor VII is dominated by the loading on the "humorous" variable. Since this variable comes on in such a strong fashion, we have labelled this factor a measure of Humor. The relationship of humor to aggressiveness is interesting, and fits rather well what we know about the "practical joker".

Factor IX seems to be a measure of Social Leadership. It is closely related to Bales (1950) concept of group maintenance. Supervisors who score high on this factor are concerned about developing strong interpersonal relationships. They have a high need to be social leaders in their work groups.

As we had hoped, the above analysis opens up many interesting avenues for further research. The factors not only seem to have a great deal of face validity, but also agree well with other research on attitudes toward supervisory behavior. Some of the earlier studies isolated two or three underlying factors; however, different researchers have suggested different factors. Our study, which was broad in scope, indicates that employee perceptions of supervisors may be more complex than first thought to be the case. We suggest that there are many factors which deserve consideration.

Discussion

The above study is not atypical of the factor analytic studies reported in the literature. There is, however, one major difference. Our "employee responses" were random data. We simply created trait responses for mythical employees by drawing a sample of numbers from Rand's Table of Random Normal Deviates, supplying each sample with an arbitrary variable name.

The moral of our story is that a reasonable person could easily convince himself, and others, that the output of his factor analysis were meaningful, unless he had set himself an a priori benchmark for evaluating his results. A researcher could conceivably be "making sense" of completely random data. Without such a benchmark reported, a factor analytic study could thus be adding little more than noise to the literature.

At a minimum, we suggest that some measure of factor reliability be made a publication requirement. Since statistical tests are not well developed for factor analysis, and certainly not widely used, we shall summarize three simple methods for checking factor reliability.

But, before we do this, let us see if our criticism of common uses of factor analysis has merely been "beating a dead horse". We went through recent issues of six academic journals and asked whether their papers utilizing factor analysis did or did not provide the reader with some sort of factor reliability measure. We also asked whether an attempt had been made to include some measure of "validity" as an index of the usefulness of the study.

This literature survey called for a fair amount of subjective evaluation.¹ Nevertheless, most of the cases were reasonably clear-cut, since we were not evaluating how successful the author was at establishing the reliability of his factors or the usefulness of his results. We were merely interested in whether or not he had made an attempt to do so.

Insert Table 2 about here

One wonders how many of the "delinquent" studies would have been published had the reliability of the factors had been measured and reported. The results of our preliminary survey were so impressive (or dismaying, depending on your point of view), that it seemed unnecessary for us to resort to a more rigorous literature survey in order to make our point.²

Methods of Reliability Measurement

Three rather simple approaches to measuring factor reliability will be considered -- split samples, a priori analysis, and Monte Carlo simulation.

- (a) Split Samples: The original data might be split into two (or more) random sub-samples. Separate factor analyses can then be run for each sub-set, and a comparison made of the solutions, for example, by correlating the factor scores as suggested by Burt (1948). We would recommend that the results from each sub-sample be published separately, so that "eye-ball" comparisons may be made, since the statistical tests for determining differences among factor structures are poorly developed (Harmon, 1960). In the studies that were surveyed in Table 2, the sample sizes were generally large enough relative to the number of variables evaluated, that it would have been easy for the authors to have split their samples.
- (b) A Priori Analysis: Before collecting or looking at the data, the researcher should work out, in as much detail as possible, the structure of the solution that he expects to find. He might, for example, postulate the number of factors he expects to appear, which variables should load together, relationships which should exist among factors, or what variables he expects will dominate which factors. His predictions could be based on behavioral models, previous findings reported in the literature, or merely on "well educated" hunches. Few of the studies that we examined made any attempt to develop an a priori model.

(c) Monte Carlo Simulation: In some cases, sample sizes are so small that it is not practical to split the sample. In addition, one may have very weak prior information about the underlying behavioral processes. For such cases we propose that researchers should try to simulate their results by factor analyzing suitable samples of random data. By "suitable samples" we mean sets of random data chosen to conform to the actual data in terms of sample size, number of variables, and assumed underlying distributions. The type of factor analysis should obviously also conform to that used on the actual data.

The reliability analysis would parallel our "study" above, except that it would be replicated many times in order to obtain distributions on the various factor statistics. By comparing the results based on actual data with the results from Monte Carlo simulations one could get an idea, based on sample frequencies, whether the former appeared to be "significantly" different from the latter. For example, comparisons could be made in terms of the following statistics:

- (1) Number of factors having eigenvalues greater than 1.0 ;
- (2) Average loading on most important variables in each factor ;
- (3) Percent of variance explained by a given number of factors .

If the actual results yielded fewer factors, had consistently higher loadings, and had a higher percent of variance explained than the simulation data, the investigator would gain increased support for a claim that his factor analysis provided a valid way of summarizing the data. On the other

hand, if actual results did not differ substantially from simulated results, one would have reason to question the value of the reported factor analysis. Several of the papers included in our survey reported results that, at face value, seemed no more impressive than the outcome of our random number experiment.

Measure of Usefulness

In addition to the measures of reliability mentioned above, it would be desirable if a researcher provided some measure of the usefulness of his factor analysis. Authors often claim that their study has been useful in an "exploratory sense". But how is such a statement to be interpreted? If the study merely provides the author with an "intuitive understanding", then it is not always clear what the next step should be. If there are other payoffs, say for prediction or control, then perhaps an author should somehow attempt to demonstrate how useful his factor analysis seems to be.

In order to judge the usefulness of an exploratory study it may, for example, be helpful to specify at least one "dependent" variable, the behavior of which the factor analysis was designed to help explain or predict. But in many of the reports that we read it appeared as if factor analysis had been run when the investigator finally saw no other way of massaging his data.

The Influence of the Analyst

Factor analysis, despite its apparent mathematical sophistication requires much interpretive artistry by the analyst. A researcher must make many decisions in order to reduce his factor analysis to a manageable size.

The rationale for making these decisions often remains unstated in published reports. In some cases in our survey, there was not even enough information for another researcher to be able to replicate the study.

Following are some of the decisions that face a researcher using factor analysis -- the first four of which will obviously be common to any multivariate analysis:

- (a) How should the variables be measured?
- (b) How many variables should be included in each analysis?
- (c) How large should the sample size be?
- (d) What types of base matrix should be operated on (e.g. Pearson product-moment correlation, a variance-covariance matrix, or a set of non-parametric correlations)?
- (e) What estimates should be used for the communalities (e.g. 1.0, highest r , multiple r^2 , subjective estimates)?
- (f) What method should be used for extracting factors?
- (g) How many factors should be extracted?
- (h) Should the factors be orthogonal or oblique?
- (i) What types of rotation should be performed?

In view of the diversity and scope of these methodological decisions it is not surprising that it can be difficult to evaluate the results of a factor analysis in which the bases for making these decisions remain unstated. To illustrate this we ran further Monte Carlo simulations on samples of normally distributed random deviates, considering "all other things equal" in each factor analysis, even to the extent of using the same sets of input data wherever possible. The following generalizations seem reasonable:

- (a) As the number of variables increases, the number of significant factors increases. For our data, of course, each variable is theoretically independent of the others, so this result is not surprising. Table 3 illustrates this effect.

Insert Table 3 about here

- (b) As the sample size increases, the percent of variance explained (by a given number of factors) decreases. Table 4 provides an idea as to the size of this effect.

Insert Table 4 about here

- (c) As the communality estimates become smaller, the number of factors with eigenvalues greater than 1.0 decreases. Once again, this is an obvious result since, as the values in the major diagonal approach zero the rank of the matrix approaches zero. Therefore, with sufficient patience and experimentation, one could obtain almost any number of factors he might desire. See Table 5.

Insert Table 5 about here

Conclusions

Our concern has been how to evaluate the results of factor analysis. We focussed on the questions of reliability and usefulness. Another useful and complementary approach would have been to examine the assumptions upon which the factor analytic model is based. This, however, has already been done by Harmon (1960). Guilford (1952) also provides a brief but informative summary of the factor analytic assumptions, with special reference to psychological testing.

With computers, factor analysis has become a relatively inexpensive technique. As a consequence, the number of published studies employing this method of analysis is rapidly increasing. Solomon (1960) reports that over 1000 papers on factor analysis were published between 1900 and 1960, and the rate of publication has increased steadily. We found that many of the factor analytic reports that we read had not provided adequate methodological information to enable their results to be evaluated by other researchers. To rely on the face validity of factor loadings is clearly not enough. We presented a simple illustration of how, on the basis of completely random data, one could produce factor analytic results that would seem both "reasonable" and "interesting".

The conclusion is that all factor analytic studies should provide a benchmark or measure of their reliability. Available statistical tests for testing the significance of a set of factors are, unfortunately, neither adequate nor well known. Those that exist are, in any case, not often used. We have indicated three approaches that might be used for checking on the reliability of a factor analytic result: split samples, a priori models

and Monte Carlo simulation. While arguments may be raised against the use of split samples or the use of a priori models in certain situations, there seem to be few cases in which a Monte Carlo approach would not be feasible.

Table 1

Nine Factors from Principal Components Solution

I		II		III	
<u>Variable</u>	<u>Loading</u>	<u>Variable</u>	<u>Loading</u>	<u>Variable</u>	<u>Loading</u>
Sensitive	-.72	Sincere	-.85	Formal	.84
Democratic	-.72			Fair	-.65
Responsible	.69			Humble	-.54
IV		V		VI	
Trustworthy	.80	Patient	.69	Kind	.82
Humble	.55	Shallow	.59		
		Aggressive	-.53		
VII		VIII		IX	
Humorous	-.88	Tactful	.81	Warm	.72
Aggressive	-.54			Active	.71
				Enthusiastic	.64
				Ambitious	.55

SUMMARY OF LITERATURE SURVEY

Journal	Years	Number of Studies	Number of Studies Without a Reliability Measure	Number of Studies Without Measure of Usefulness	Number of Studies With Neither Reliability Nor Useful- ness Measures
Educational and Psychological Measurement	1964,1965,1966	11	4	6	4
Journal of Applied Psychology	1963,1965	15	11	10	8
Journal of Consulting Psychology	1964,1965,1966	7	4	2	2
Journal of Marketing Research	1964,1965,1966	3	2	2	1
Journal of Personality	1964,1966	1	1	1	1
Personnel Psychology	1962,1963,1965	9	8	6	5
		—	—	—	—
Totals		46	30	27	21

Table 3

Effects of Variation in Number of Variables

For Principal Components -- n = 50

<u>Number of Variables</u>	<u>Number of Factors with Eigenvalues > 1.0</u>
20	8
40	16

Table 4

Effects of Variation in Sample Size

For Principal Components -- 20 variables 8 factors

<u>n</u>	<u>Percent of Variance Explained</u>
30	72.8
50	65.1
100	58.0

Table 5

Effects of Variation in Communalities Estimates

For Principal Components -- 20 variables; n = 50

<u>Communalities Estimates</u>	<u>Number of Factors with Eigenvalues > 1.0</u>
1.0	8
multiple r^2	4
highest r in row	3

FOOTNOTES

1. Factor analytic reports excluded from the survey were those studies that explicitly set out to test, or refine, the reliability of some previously reported, factor-derived scaling of items (like the MMPI scale), or to replicate, or test, the predictive validity of some previously reported set of factors.

2. The articles surveyed are not listed in the bibliography. However, anyone wishing to know more about which particular articles were chosen, or how they were coded, is invited to write to either author.

BIBLIOGRAPHY

- Adorno, T.W., Else Frenkel-Brunswick, D.J. Levinson and R.N. Sanford, The Authoritarian Personality, New York: Harper, 1950.
- Bales, R.F., Interaction Process Analysis: A Method for the Study of Small Groups, Cambridge, Massachusetts: Addison-Wesley, 1950.
- Burt, C., "The Factorial Study of Temperamental Traits", British Journal of Psychology, Statistical Section, 1 (1948), 178-203.
- Dixon, W.J. (ed.), Biomedical Computer Programs, Revised Sept. 1, 1965, U. of California at Los Angeles.
- Fiedler, F.E., "The Leader's Psychological Distance and Group Effectiveness," in D. Cartwright and A. Zander (eds.), Group Dynamics, Chicago: Row Peterson, 1960.
- Guilford, J.P., "When Not to Factor Analyze", Psychological Bulletin, 48 (1952), 26-37.
- Hertzberg, F., B. Mausner and B.B. Snyderman, The Motivation to Work, New York: John Wiley, 1959.
- Harmon, H., Modern Factor Analysis, U. of Chicago Press, 1960.
- Kaiser, H.F., "The Application of Electronic Computers to Factor Analysis", Educational and Psychological Measurement, 20 (1960), 141-151.
- Milgram, S., "Behavioral Study of Obedience", in W.G. Bennis et.al. (eds.), Interpersonal Dynamics, Homewood, Illinois: Dorsey (1964) 108-121.
- Rand Corporation, A Million Random Digits with 100,000 Normal Deviates, Free Press of Glencoe, 1955.
- Solomon, H., "A Survey of Mathematical Models in Factor Analysis", in H. Solomon (ed.), Mathematical Thinking in the Measurement of Behavior, Free Press of Glencoe, 1960.

28.23 '71

BAGGAGE

Date Due

APR 12 '78

NOV 7 '78

DEC 11 '78

NOV 23 '78

DEC 11 '78

NOV 23 '78

Lib-26-67

MIT LIBRARIES



3 9080 003 870 349

254-67

MIT LIBRARIES



3 9080 003 901 342

255-67

MIT LIBRARIES



3 9080 003 870 331

256-67

MIT LIBRARIES



3 9080 003 901 441

258-67

MIT LIBRARIES



3 9080 003 870 364

254-67

MIT LIBRARIES



3 9080 003 870 323

259-67a

MIT LIBRARIES



3 9080 003 870 570

260-67

MIT LIBRARIES



3 9080 003 870 588

261-67

MIT LIBRARIES



3 9080 003 870 620

262-67

MIT LIBRARIES



3 9080 003 901 565

263-67

